

Standards and Assessments Division

**New England Common Assessment Program:
October 2005 Administration Review and Evaluation**

EXECUTIVE SUMMARY

The New England Common Assessment Program (NECAP), a state level testing program developed collaboratively by the Vermont, Rhode Island and New Hampshire Departments of Education, became operational in October 2005. More than 216,000 students in grades 3 through 8, with approximately 20% representing 322 Vermont schools, were assessed in reading and math. Students in grades 5 and 8 were also assessed in writing.

The purpose of this report is to summarize feedback that was gathered during and after testing, and to offer recommendations for improving future NECAP administrations. The data represent the opinions of more than 2000 educators, as well as some students, who completed questionnaires, participated in debriefing sessions, or sent letters and e-mails to the Department. Because many aspects of the testing program were designed in response to federal regulations, and therefore are not subject to change, the discussion and recommendations are focused on actions that do fall within the purview of the Department to modify and improve.

Findings:

The overall development and design of NECAP appears to be sound. Comments suggested that the tests fit well with classroom priorities, are aligned with the state's Grade Level Expectations (GLE), are free from bias, and are generally an improvement over the state's previous state assessment program. A number of responses suggested that standardized testing in third grade, a federal requirement, is developmentally inappropriate. However, comments on how the youngest students responded to testing were notably diverse; some teachers reporting that students were in tears and others writing that students enjoyed the testing and wanted to do more.

Feedback on testing materials was also generally positive, with at least 75% of questionnaire respondents rating coordinator and administrator manuals, accommodation guides, practice tests and student preparation materials as good to excellent. Almost 15% indicated that they were not familiar with practice tests, 21% were not familiar with the test preparation documents, and 13% were not familiar with the accommodation guide. Suggestions were offered for improving the test booklets, student testing resources and administration scripts.

Although reports on the average length of test sessions varied widely, from less than 30 minutes to more than an hour, a half dozen comments were received suggesting that testing resulted in a significant loss of instructional time. However, none of the suggested time-saving alternatives proved to be feasible, either because the option would fail to meet federal requirements or because it would negatively impact on features of the testing program that improve accessibility for students with special assessment needs.

Various logistical aspects of the testing program, including training, support services and distribution procedures, received very satisfactory ratings from test coordinators. However, because federal testing requirements have doubled or tripled the number of grades that need to be assessed in many schools, principals and test coordinators frequently expressed concerns about

time demands, strained resources, space problems and scheduling conflicts that characterized their experience with the assessment..

Finally, feedback indicated that accessibility options for students with special assessment needs were satisfactory, with 85% of test administrators rating the NECAP accommodation guide as good to excellent. Thirteen percent indicated that they were not familiar with the accommodations guide. The out of level testing option, which was not an original component of the NECAP design and was available only for Vermont students, received less favorable ratings because of logistical and procedural problems with authorization, ordering and distribution.

Recommended Actions:

Working with the testing contractor, the Department will continue to gather information that may help explain why some Vermont schools had less favorable experiences with NECAP than most, given that all schools used the same tests, had access to the same training and support, and worked from the same set of policies, logistics and administration procedures. The Department will continue to revise assessment manuals and materials based on feedback from the field. Opportunities for schools to share strategies and effective practices will be provided through test administration and test interpretation workshops. New accessibility options will be explored, and logistical procedures for using them will be refined. New or improved uses for technology will be pursued, including use of the Department's NECAP webpage to post new materials and guidelines.

Because schools are ultimately responsible for preparing the students, implementing policies and procedures, and giving the tests, they play a critical role in the success of NECAP. Every school should have an assessment plan and at least one test coordinator. They should send two representatives to administration workshops. They should make sure staff are trained and have access to manuals and support materials. They should make sure students are fully prepared for testing.

There is some evidence in the data gathered during this review to indicate that not every school participated fully with the supports that were available to best administer the tests. This may provide a partial explanation for why some schools had a less positive experience with NECAP than others. NECAP was developed using the best assessment models and strategies currently available, but the program can only succeed if it can be administered accurately and efficiently by the schools. Through the collaborative efforts of the testing contractor, the Department, and the schools, state assessment can be worth the effort, with the payoff measured in terms of improved student achievement.

Standards and Assessments Division

**NEW ENGLAND COMMON ASSESSMENT PROGRAM:
OCTOBER 2005 ADMINISTRATION REVIEW AND EVALUATION**

DECEMBER 30, 2005

The reading, math and writing tests that make up the New England Common Assessment Program (NECAP) were administered to students in Vermont, New Hampshire and Rhode Island for the first time in October, 2005. More than 216,000 students in 1126 schools participated, including 42,276 Vermont students in 322 Vermont schools. Math and reading were assessed in grades three through eight, and writing was assessed in grades five and eight. The purpose of this report is to summarize feedback gathered during and after the October NECAP test administration and to make recommendations for improving future NECAP administrations.

NECAP emerged from the comprehensive and detailed testing requirements of the federal No Child Left Behind Act (NCLBA). The Department acknowledges that there are diverse perspectives on NCLBA. Addressing those perspectives is not the purpose of this report. When Vermont accepted NCLBA funds it also accepted the reality that NCLBA sets relatively narrow parameters for the design and administration of state assessments, and that it also creates a significant amount of new work for the Department, schools and students. The review and recommendations in this report are based on the assumption that accountability assessment can be a valuable part of school improvement. The focus is on the aspects of the testing program that are in the purview of the Department to modify and improve within the context set by NCLBA.

It should also be noted that NECAP was developed through a collaborative effort of the Vermont, Rhode Island and New Hampshire State Departments of Education. The governance structure for the project provides each state with an equal vote on all decisions, and requires mutual consent by all three states on all components of the program. As result, some of the recommendations included in this report can be implemented only with the approval of Rhode Island and New Hampshire. The report indicates whenever that is the case.

The following data sources were used to prepare this report:

- **NECAP Test Coordinators' Questionnaire**

Seven hundred seventy-seven Vermont principals or their designees responded to some or all of this survey which covered a range of topics related to test administration procedures, materials and support services.

- **NECAP Test Administrators' Survey**

Two thousand twenty-six educators who participated in the administration of the NECAP tests responded to questions about the quality of manuals and practice tests, administration procedures and length of test sections.

- **NECAP Debriefing Summaries from Mathematics Network Meetings**

Approximately 150 third and fourth grade teachers discussed a variety of topics related to the NECAP administration, including student reactions, preparation for testing, test directions for teachers and students, test resources and materials, and test design.

- **E-Mails and Letters Received by the Department of Education**

After the completion of the October 2005 NECAP administration, Commissioner Cate requested participating schools to send the Department feedback and recommendations. A total of 53 e-mails from school personnel were received. In addition, 22 students sent comments by mail or e-mail.

Results are presented in five sections: Test Design, Materials, Administration, Logistics and Student Accessibility. At the end of each section specific actions are recommended for the NECAP assessment contractor, Measured Progress, and for the Department of Education. Suggestions for how schools can help improve NECAP administrations are also offered.

REVIEW, EVALUATION AND RECOMMENDED ACTIONS

Test Design:

The test design factors that are of particular interest when evaluating the first administration of an assessment include the quality of individual test items, the range and balance of content emphasis across test items, the range and balance of cognitive complexity (also known as depth of knowledge) across test items, and the alignment between test items and the Grade Level Expectations (GLE) they are designed to assess.

To a great extent, the feedback that is NOT received about test design can be as telling as the feedback that is. As a result, it is very encouraging that no “bad items” (e.g., ambiguous questions or answer choices or items that reflect inappropriate pedagogy) were cited in the feedback. Of the more than 1000 items that make up the tests, only one suggestion of misalignment was received – a math item that used the word “most” when the corollary GLE uses the word “more.” The item will be replaced. An entire 8th grade class from Vermont’s Northeast Kingdom wrote to suggest that a writing prompt referencing cell phones was regionally biased because cell phones don’t work in their part of the state. Item statistics do not substantiate the students’ concern. Two respondents suggested that a particular item on the 4th grade math test was designed to trick students – “trick items” are not part of the NECAP design – but the item was, in fact, found to be consistent with the 3rd grade GLE.

A few teachers suggested that the test would have been more appropriate for their students at the end of the school year rather than the beginning. In contrast, a Math Network Leader’s summary of a NECAP Debriefing session suggested that that test accurately reflected the GLE, but local curriculum did not. She wrote, “Teachers found items, skills and processes that have been missed at a grade level and/or needs more teaching emphasis.” Readers should note that every NECAP item was reviewed for GLE alignment by teams of teachers from Vermont, Rhode Island and New Hampshire. Items that were not correctly aligned were either edited or removed from the item pool. All items were subsequently field tested and were removed from the item pool if item

statistics indicated that the item was not appropriate for the specific grade level (i.e., more than 20% of grade level students got the item incorrect or more than 90% got the item correct).

The single greatest concern about the NECAP test design can be summarized by the following comment from a principal: “Third Grade students were not prepared for this kind of test. It’s hard for this age to stay focused on tasks independently.” This comment was echoed by several primary level teachers who suggested that standardized testing, no matter how well designed, is not developmentally appropriate for third graders. However, this was not a universal concern. In fact, the diversity of comments on how the youngest test takers responded to NECAP testing was striking, from reports of students who were “overwhelmed” and “in tears” to comments such as “students seemed to enjoy the challenge of standardized testing and felt that they did well” and “they liked it; they wanted to do more.”

One test coordinator suggested that the range of student reactions might be explained by how well students were prepared for testing, commenting that practice tests were “very useful; kids were more comfortable after using them.” Another suggested that it was academic preparation that was the made the difference, writing “some students were overwhelmed by NECAP but they’re the ones that are overwhelmed by everything.” Testing at the third grade level is an NCLBA requirement. Practice tests and annual release of a percentage of test items are strategies that were built in to NECAP to help prepare all students, but particularly young students, for testing.

Overall, feedback concerning test design was positive. One administrator, for example, wrote, “In general, our faculty felt the NECAP items on the test were appropriate.” Another wrote, “My sense from a cursory examination is that the test questions are more thoughtful than many in the past.” In particular, NECAP was rated favorably in comparison to Vermont’s previous assessment, the New Standards Exam (e.g., “NECAP is a better fit with the classroom than NSRE” and “has much better questions than NSRE”). Network meeting attendees suggested that they needed to learn more about NECAP test design features such as depth of knowledge, aligning assessments with GLE, and universal design so that they can apply those concepts to their classroom assessments.

What will Measured Progress do?

1. The math item that was found to be misaligned with GLE has already been pulled from the test. It will be revised and field-tested a second time.
2. The “cell phone” writing prompt will not be used in future test administrations. The student feedback about this prompt has been integrated into the training packet for Bias-Sensitivity reviewers.

What will the Department do?

1. The Department will continue to embed information on GLE alignment, balance of representation, depth of knowledge and universal design into Vermont Professional Development Network offerings on local assessments, as well as regional reporting workshops and Test Administration workshops.
2. The Department will also offer additional guidance on how practice tests, released items and other test preparation strategies can be used, particularly with 3rd graders for whom NECAP is their first experience with standardized testing.
3. The Department will continue to investigate reasons why students in different schools responded differently to the assessment experiences. Promising strategies for preparing

students and improving test administration will be discussed at future test administration workshops, and results will be disseminated using the Department's NECAP webpage.

4. The Department will continue to recruit a broad range of teachers to be involved in item review and bias-sensitivity review.

What can schools do?

1. Make sure that at least two staff members attend the Principal/Test Coordinator Workshops that will be scheduled during the month prior to the October 2006 NECAP assessment.
2. Use practice tests and released items as part of the beginning of the year review process. Discuss test-taking strategies with students. Consider having 4th graders talk to 3rd graders about testing. Overall, when preparing students for testing, try to find the critical balance between encouraging students to do their best without creating undue stress.
3. Make sure teachers are acquainted with the NECAP Accommodations Guide so that accommodations can be planned and used effectively.
4. Continue to align curriculum and classroom instruction with the Vermont Grade Level Expectations so that students can be prepared academically for assessment.

Test Materials:

The review of testing materials covers two areas: supporting materials such as manuals, guides and practice tests, and the actual testing materials such as test booklets, student response booklets and student testing resources.

Eighty-two percent of the more than 2000 teachers who completed the Test Administrator Questionnaire rated the quality of information and instructions in the *Test Administrator Manuals* as good to excellent. Only 1.8 percent said it was inadequate. Seventy-Six percent rated the quality of student instruction scripts as good to excellent, with only 3% indicating that the scripts were inadequate. Principals and test coordinators also gave their manual high ratings, with 82% indicating that the *Principal/Test Coordinator Manual* was Good or Very Good, and only 1.6% rating it as poor.

Support materials also received good marks. The practice tests were rated good to excellent by 75.9% of respondents and a test preparation document called *Preparing Students for NECAP: Tips for Teachers to Share with Students* was rated good to excellent by 74.3%. Of the respondents who did not rate these documents as good or excellent, 14.7% indicated that they were not familiar with the practice tests and 21% indicated that they were not familiar with the test preparation document.

Comments on the testing materials were generally positive, with test administrators describing them as "sufficient" and "working well." Several teachers requested that in the future the grades 5 through 8 tests use an integrated test/answer booklet so that students don't need to put their answers or show their work in a separate student response booklet. It was also noted that the "stop signs" at the end of test sessions needed to be enlarged, and a blank sheet should be placed between sections, so that students who work quickly will be less likely to continue working into the next section of the test.

With respect to the student testing resources, several teachers indicated that the pattern blocks, which needed to be "popped out" of a perforated sheet of oak tag paper, were tedious for

teachers to prepare but too difficult for students to manage. Finally, several teachers were concerned that some students were confused by the testing resources, particularly the calculators, because the test directions didn't make it clear that none of the test items specifically require that the resources and materials be used.

What will Measured Progress do?

1. Continue to update and improve manuals in response to feedback from test administrators and test coordinators.
2. Add information to the student scripts to clarify that student testing resources are to be used at the student's discretion and that no items on the test specifically require the use of a calculator.
3. Work with the Department to prepare test interpretation materials, which will be disseminated via regional reporting workshops and website postings, including materials that will help schools make effective use of the release items as part of student test preparation for future NECAP administrations.

What will the Department do?

1. Continue to offer workshops, materials and web postings that will help teachers and administrators become aware of accommodation guides, practice tests, student preparation tips, and any new materials that are developed in the future.
2. Investigate the possibility of eliminating the separate student response booklet in grades 5 through 8 to reduce the amount of material students need to have on their desks during testing, and to eliminate the need to put answers and work in a separate booklet. (Note: this action will require agreement from New Hampshire and Rhode Island, and may have cost implications).
3. Investigate the possibility of having Measured Progress pre-package manipulatives, or at a minimum, increase perforations to make it easier for students to prepare the manipulatives. (Note: this action will require agreement from New Hampshire and Rhode Island, and may have cost implications).

What can schools do?

1. Make sure that all teachers have access to and are using all currently available NECAP resources such as accommodation guides, practice tests and student preparation tips, as well as any other NECAP support materials that may become available in the future. The Department believes that use of support materials, particularly practice tests and student testing tips, will maximize the testing experience for most students, and will prevent many of the problems that were reported after the October 2005 testing. However, whether or not the support materials are used is entirely up to the schools.

Test Administration Procedures and Timelines:

Test administration, particularly the amount of time that is devoted to testing, raised the greatest number of concerns about NECAP. Seven of the 53 e-mails received by the Department included comments similar to the following:

Way too much time. For example, in most of our classes some needed/used 90 minutes. This meant grades 5 and 8 took 12 hours of testing time. This does not include directions and practice. That's over 3 days of instructional time.

Reports on the amount of time that was needed to complete testing varied widely. In contrast to the e-mail reprinted above, 72% to 78% of the Test Administrator Questionnaire respondents indicated that most students were able to complete test sessions in less an hour, with mathematics appearing to take slightly longer than reading or writing. Only 5% to 7% indicated that sessions lasted 76 to 90 minutes, again the difference depending on the content areas that was tested.

It should be noted that the NECAP tests don't have specific time limits. Test sessions were designed so that most students will be able to finish in less than 45 minutes. However, at the test administrator's discretion, students who are working productively can be allowed up to 45 additional minutes for a total of 90. Students who finish early are permitted to read a book or work quietly. This means that many students in a class may finish a test session in an hour or less, but that 90 minutes must be scheduled for each session to accommodate the few students who may need more time. It is important to be clear on this point. While it is true that up to 90 minutes of instructional time may be lost for each session of NECAP testing, it does not mean that students are routinely subjected to a 90 minute test.

Several of the e-mails included suggestions related to reducing the overall length of the test – either cutting the test down to two sessions or decreasing the overall number of test items. A principal noted that “all three testing sessions were quite similar,” and asked “what do we learn by testing for three sessions that we didn't learn in the first two sessions?”

According to NECAP's test developers, the answer to the principal's question has to do with the technical adequacy of the test. First, it should be noted that NCLBA requires states to assess the breadth and depth of state reading and mathematics standards. Prior to the development of the NECAP tests, significant attention was paid to test length because of concerns raised about this issue by the field and the Vermont State Board of Education. As result, NECAP test specifications were based on the minimum number of items needed to meet NCLBA requirements and to maintain test validity and reliability – approximately 50 to 60 item points for each content area at each grade level.

Because it would not be feasible to reduce the number of test items, an alternative option for cutting NECAP administration time might be to make the tests timed, that is, to restrict the length of each test session to exactly 45 minutes with no teacher discretion option for extending time when students are working productively. An individualized plan listing extended time accommodations and testing in a separate location would be required for students who need more than 45 minutes. There are, of course, a number of problems with this option. First, it would add an element of cognitive complexity to test items that is not intended in the GLE, changing what are supposed to be power tasks to fluency tasks. It would also prevent teachers from allowing larger groups of students to have a few minutes of additional time when warranted, as well as placing an undue burden on school teams to identify, plan and find staff and space for all students who may need more than 45 minutes. Finally, it would represent a major departure from the principles of universal design that are a key feature of the NECAP tests.

Two e-mails suggested that the impact of testing on instructional time could be reduced by eliminating the 5th and 8th grade writing tests, since writing assessment is not an NCLB requirement. During discussions about testing prior to NECAP development, many Vermont educators expressed interest in continuing the State's practice of assessing writing at the state level. However, since it was not an NCLBA requirement, the consensus was to maintain writing assessment at two grades. NECAP was designed based on this input from the field.

Several comments were received about the test directions and administration scripts. A few test administrators suggested that some of the language in the student scripts was awkward or that it was "...stated in a different way than the daily classroom language." Some were also concerned that teachers were not allowed to paraphrase directions. Two test administrators objected to the prohibition on reading complete math items to students (only single words were allowed to be read).

Ten of the e-mails received by the Department concerned the October testing window. Two supported fall testing, but suggested expanding the test window to a full month. Six expressed concern that this year's testing window coincided with Yom Kippur and Rosh Hashanah, and that some schools scheduled testing on those dates. Readers should note that the Department sets a broad assessment window – October 3rd through 26th in 2005 - but the schools select the days when the 6- 8 sessions of testing will actually occur. The remainder of the e-mails, as well as a few comments from the 3rd grade math teachers, suggested several reasons why October was not an ideal time to administer the tests. When NECAP was developed, several scheduling options were considered. Though none proved to be perfect, fall testing was the best option with respect to meeting NCLBA requirements and getting assessment results back to schools at a time when the data would be most useful.

What will Measured Progress do?

1. Continue to revisit the student scripts each year prior to testing. Involve teachers in reviewing language used in directions so that they are clear, concise and are consistent with language used in classroom instruction.

What will the Department do?

1. Investigate the possibility of expanding the parameters of support that can be provided to students during assessment, particularly related to clarifying instructions and reading items that don't measure reading. If it is not possible to permit increased interaction between students and test administrators, provide a clear rationale for these prohibitions. (Note: this action will require agreement from New Hampshire and Rhode Island).
2. Encourage schools to consider religious holidays when developing their assessment schedule. Include that suggestion in the Test Coordinator's manual and the test administration workshops.

What can schools do?

1. When it is clear that a small number of students may delay completion of a test session for the class, plan extended time accommodations in a separate setting for those students so lost instructional time is minimized for the majority of classmates. (Please Note: NECAP does not restrict accommodations to just special education students).

Logistics:

The results of the Principal/Test Coordinator Questionnaire indicate a high level of satisfaction with the information and support provided by the Department and Measured Progress before, during and after testing. Eighty-Five percent of the principals and test coordinators indicated that the information provided to schools prior to test administration (mailings, website postings) was good or very good, with only 1.1 percent indicating that it was poor. Logistics associated with the receipt of materials was rated good or very good by 89.2% of respondents, with 1.7% selecting a poor rating, and logistics related to the return of test materials was rated good or very

good by 72.1% of respondents, with 5% selecting the poor rating. Eighty-three percent of the principals and test coordinators indicated that the test administration workshops held during the month prior to assessment were somewhat helpful or very helpful, with 12.4% indicating that they had not attended the administration workshops. The Measured Progress Service Center was rated adequate to good by 69% of the respondents, with 26.7% indicating that they didn't use the Service Center.

The time required of principals to plan for the assessment and to process the testing materials was a primary concern expressed in several of the letters and e-mails. One principal wrote, "I estimate that I personally spent more than 50 hours preparing and coordinating the testing." Although there was no evidence that the NECAP tests were any more or less difficult to process than Vermont's previous assessments, the NCLBA requirement to assess students in grades 3 through 8 and once in high school has added four grades to Vermont's state assessment program. In many schools this requirement has doubled or tripled the number of tests that needed to be unpacked, counted, labeled, distributed, collected, recounted, sorted, packed and returned to the testing contractor. Also doubled or tripled was the need for qualified staff to proctor test sessions and provide individual student accommodations, the need for space to administer the assessments, and the number of random conflicts, challenges and surprises that come with any large scale assessment. Many Vermont schools have designated two or more assessment coordinators, but in those schools where all the responsibilities and challenges fell on one individual, the task was formidable with few options to make it less so.

What will Measured Progress do?

1. Clarify the procedures and directions for the On-Line Principal certification system. Offer specialized support for individuals who are not "web-savvy" to avoid creating a burden with technology that is meant to save time and effort.
2. Make sure the on-line verification system has sufficient capacity to handle the surge of users during the final days of the assessment window.

What will the Department do?

1. Investigate the possibility of extending the testing window to at least one full month, allowing additional time for principals and test coordinators to prepare materials for shipping. (Note: This action will require agreement from Rhode Island and New Hampshire, and may have cost implications).
2. During the test administration workshops, discuss strategies for streamlining assessment logistics. Provide participants with an opportunity to share effective strategies and practices. Disseminate a "tips list" through postings on the Department's NECAP webpage. Identify and recruit test coordinators who had successful experiences with the NECAP administration who would be willing to serve as a resource to schools seeking strategies to address problems with logistics or other test administration concerns.
3. Explore ways to maximize the number of students who have pre-printed labels for their test booklets.
4. Work with Measured Progress, New Hampshire and Rhode Island to explore and expand computer and web-based options for streamlining assessment logistics, including a mechanism for schools to print student labels for new students. (Note: This action will require agreement from Rhode Island and New Hampshire, and may have cost implications).

5. Post administration manuals and coordinators manuals on the Department website at least two weeks prior to the test administration workshops so that test coordinators can review them and formulate questions before attending Test Administration workshops.

What can schools do?

1. Develop an effective test administration plan and schedule, including methods appropriate to the school for sharing the many responsibilities involved in preparing, administering and returning the tests.
2. Make sure that each school is represented by one and preferably two designated test coordinators at the test administration workshops. These individuals should be committed to reading the test coordinators and test administrators manuals prior to the meetings so that they can formulate key questions.

Accessibility:

Accessibility refers to elements and option in a testing program that make it possible for all but a tiny percentage of students to participate in assessment. Maximizing accessibility has been a major goal for the Department since the inception of its state testing program. Consistent with that goal, the 2005 NECAP provided a variety of accessibility options, including an extensive list of assessment accommodations that were available to all students and “universal design” features that facilitates participation for students with an array of special assessment needs. In addition, and despite opposition from the US DOE to Out-of-Level Testing (OLT), Vermont was the only one of the three NECAP states that permitted schools to assess some students at their developmental level rather than their current grade placement.

Nearly 85% of test administrators who completed the questionnaire rated the NECAP *Accommodations Guidelines and Procedures* document as good to excellent, with only 2.7% indicating that it was inadequate. Unfortunately, 12.6% of the respondents reported that they were unfamiliar with the *Accommodations Guide*. Problems and concerns regarding accommodations included inadequate staff and space to provide accommodations, and the time required to implement accommodations for some students. In particular, these were concerns mentioned frequently by smaller schools. Several test administrators reported that the practice tests were useful for preparing students with special assessment needs.

Several respondents reported logistical problems with the OLT alternate assessment option. OLT was not a component of the original NECAP test design. It was made available for Vermont students when it became apparent that the New England Compact’s Enhanced Assessment Grant would not produce a viable alternative for OLT in time for the October 2005 NECAP administration. As result, it was necessary for the Department and Measured Progress to retrofit procedures for ordering, approving and shipping OLTs into an existing distribution system that wasn’t designed to accommodate the additional assessment option. According to e-mails, this lead to frustration for some test coordinators and a fair amount of inefficiency. Several test coordinators reported needing to make multiple calls to get OLTs approved for new students, in some cases resulting in materials that didn’t reach the school in time for the first day of testing.

With the recent US DOE publication of proposed regulations for assessment of students with disabilities, future plans for how Vermont will meet the needs of students who have traditionally accessed OLTs are not clear at this time. New options and methods will need to be discussed and

any logistics that caused problems in 2005 will need to be addressed in the plans for the 2006 administration of NECAP.

What will Measured Progress do?

1. Work with the Department to identify possible alternatives to OLT.
2. If OLT continues to be an Alternate Assessment option for Vermont, or if new options for these students are identified, work with the Department to develop more efficient procedures for ordering, approving and shipping these assessments.

What will the Department do?

1. Make sure alternate assessment staff are in the office during the two weeks prior to testing.
2. Find ways to streamline the approval and ordering process for OLTs, or any new alternate assessments that take the place of the OLT option.
3. If OLT, or any other alternate assessment option, continues to be available exclusively for Vermont students, ask Measured Progress to designate one or two service center representatives to handle all related calls. (Note: this action may have fiscal implications).
4. Make sure alternate assessment verification reports are sent to schools at least one month prior to assessment. Using the best available information, send the reports for transitioning students to both sending and receiving schools.

What can schools do?

1. Assign at least one staff member per district to function as an alternate assessment mentor. The alternate assessment mentor will serve as a contact for Department staff and will be responsible for ensuring that all staff members are aware of dates, policies and materials related to accommodations and alternate assessments.
2. Make sure to submit Documentation of Eligibility forms to the Department prior to the May 15th deadlines.
3. When student transition from one grade or one school to another, make sure a discussion about assessment participation options is part of the transition, particularly if a student has been identified for an OLT.

SUMMARY AND CONCLUSIONS

In the first section of this report it was noted that the best measure of a new assessment program is often the feedback that is NOT received. Thus it is noteworthy that none of the feedback reviewed for this report suggested that the NECAP tests are poorly designed or that they miss the mark on instructional relevance. No “fatal flaws” were uncovered. A number of problems and concerns were reported, although none of them appeared to be universal. In fact, the range of opinions on the assessment was remarkably diverse.

Although reports indicate that most schools had a satisfactory experience with NECAP, a few expressed concerns about time, logistics and student reactions to the testing. The diversity of feedback may speak to the greatest challenge that needs to be addressed for NECAP to succeed. Every Vermont school that participated in NECAP used the same tests, had access to the same training and support, and worked from the same set of policies, logistics, and administration procedures. Yet according to the data gathered for this review, the quality of schools’ experiences with the testing program varied widely. Working together, the Department, Measured Progress and Vermont educators need to determine what made the difference. We

need to work together to identify and disseminate strategies and practices that led to a positive experience in many schools. We also need to identify and correct factors that were present in some schools, but not others, that led to problems and concerns. Some of the remedies have already been identified. Certainly we will discover others as we move ahead through the first full implementation year of NECAP.

There is no escaping the reality that because of federal testing requirements, schools are now administering more tests than ever - more perhaps than they feel are necessary - dramatically increasing demands on their time, space and human resources. However, it is the reality in which any actions to improve NECAP must be taken. The Department and Measured Progress will make a good faith effort to take the actions recommended in this report, and will continue to seek feedback in an on-going process of evaluation and improvement. Schools need to be full partners in this endeavor

Schools play an important role in making NECAP work because schools are ultimately responsible for preparing the students, implementing the policies and procedures, and giving the tests. Every school should have an assessment plan to ensure that teachers and students are prepared, that resources are used efficiently, and that conflicts are avoided. Every school should designate at least one test coordinator, and preferably two, who will attend administration workshops, take the information back to colleagues, and take primary responsibility for preparing and implementing the school's assessment plan. Test coordinators and test administrators should be fully apprised of and follow administration policies and procedures. Announcements and updates from the Department and Measured Progress should be read and disseminated, and appropriate actions should be taken. Support materials and practice tests should be accessed and used. There is evidence in the data that not every school participated fully with the supports available to best administer the tests. The Department and Measured Progress have attempted to create systems, materials and resources to minimize the impact of assessment on schools, but they can only be effective if schools use them as intended.

In many ways, NECAP has no parallel in the other 47 states, particularly with respect to teacher involvement in the test development process. Teams of Vermont, Rhode Island and New Hampshire teachers:

- Developed the Grade Level Expectations and helped determine distribution of emphasis – how items and points should be allocated to specific GLE clusters within and across test grade levels;
- Reviewed every test item, suggesting revisions when necessary, to ensure that each was sufficiently aligned with the GLE;
- Reviewed every reading passage and every test item to ensure that each was free from racial, religious, ethnic, gender or regional bias, rejecting those that weren't;
- Participated in setting the tests' proficiency standards.

It is also worth noting that NECAP is one of the first state assessment programs to:

- Employ Norman Webb's Depth of Knowledge procedures in item development, ensuring that students are challenged to demonstrate their learning along a continuum of cognitive complexity;
- Use principles of universal design so that the tests will be maximally accessible to all but a very few students
- Create an interstate collaboration model for how small states can work together to do big things.

NECAP was developed using the best models and strategies currently available, with on-going support and oversight from a technical advisory panel that includes some of the foremost assessment experts in the country. However, the program can only succeed if it can be administered accurately and efficiently by our schools. Through the collaborative efforts of the testing contractor, the Department, and the schools, state assessment can be worth the effort, with the payoff measured in terms of improved student achievement.